

M. Sc (Information Technology)		Semester – I	
Course Name: Data Science		Course Code: PSIT102	
Periods per week 1 Period is 60 minutes	Lectures	4	
	Credits	4	
		Hours	Marks
Evaluation System	Theory Examination	2½	60
	Theory Internal	--	40

Objectives	<input type="checkbox"/> Develop in depth understanding of the key technologies in data science and business analytics: data mining, machine learning, visualization techniques, predictive modeling, and statistics. <input type="checkbox"/> Practice problem analysis and decision-making. <input type="checkbox"/> Gain practical, hands-on experience with statistics programming languages and big data tools through coursework and applied research experiences.
-------------------	--

Pre requisites	Basic understanding of statistics
-----------------------	-----------------------------------

Unit	Details	Lectures
I	Data Science Technology Stack: Rapid Information Factory Ecosystem, Data Science Storage Tools, Data Lake, Data Vault, Data Warehouse Bus Matrix, Data Science Processing Tools ,Spark, Mesos, Akka , Cassandra, Kafka, Elastic Search, R ,Scala, Python, MQTT, The Future Layered Framework: Definition of Data Science Framework, Cross-Industry Standard Process for Data Mining (CRISP-DM), Homogeneous Ontology for Recursive Uniform Schema, The Top Layers of a Layered Framework, Layered Framework for High-Level Data Science and Engineering Business Layer: Business Layer, Engineering a Practical Business Layer Utility Layer: Basic Utility Design, Engineering a Practical Utility Layer	12
II	Three Management Layers: Operational Management Layer, Processing-Stream Definition and Management, Audit, Balance, and Control Layer, Balance, Control, Yoke Solution, Cause-and-Effect, Analysis System, Functional Layer, Data Science Process Retrieve Superstep : Data Lakes, Data Swamps, Training the Trainer Model, Understanding the Business Dynamics of the Data Lake, Actionable Business Knowledge from Data Lakes, Engineering a Practical Retrieve Superstep, Connecting to Other Data Sources,	12
III	Assess Superstep: Assess Superstep, Errors, Analysis of Data, Practical Actions, Engineering a Practical Assess Superstep,	12

IV	<p>Process Superstep : Data Vault, Time-Person-Object-Location-Event Data Vault, Data Science Process, Data Science,</p> <p>Transform Superstep : Transform Superstep, Building a Data Warehouse, Transforming with Data Science, Hypothesis Testing, Overfitting and Underfitting, Precision-Recall, Cross-Validation Test.</p>	12
V	<p>Transform Superstep: Univariate Analysis, Bivariate Analysis, Multivariate Analysis, Linear Regression, Logistic Regression, Clustering Techniques, ANOVA, Principal Component Analysis (PCA), Decision Trees, Support Vector Machines, Networks, Clusters, and Grids, Data Mining, Pattern Recognition, Machine Learning, Bagging Data, Random Forests, Computer Vision (CV) , Natural Language Processing (NLP), Neural Networks, TensorFlow.</p> <p>Organize and Report Supersteps : Organize Superstep, Report Superstep, Graphics, Pictures, Showing the Difference</p>	12

Books and References:					
Sr. No.	Title	Author/s	Publisher	Edition	Year
1.	Practical Data Science	Andreas François Vermeulen	Apress		2018
2.	Principles of Data Science	Sinan Ozdemir	PACKT		2016
3.	Data Science from Scratch	Joel Grus	O'Reilly		2015
4.	Data Science from Scratch first Principle in python	Joel Grus	Shroff Publishers		2017
5.	Experimental Design in Data science with Least Resources	N C Das	Shroff Publishers		2018